SVMs via Convex Hulls

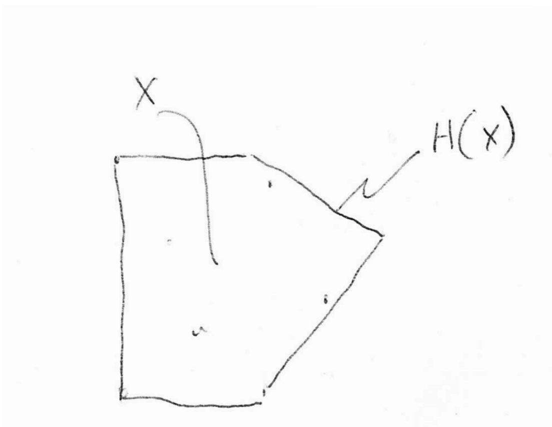
Instead of developing SVMs via Langrangian optimization theory we can develop SVM using convex hulls.

Convex Hulls

Let $X = \{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_l\} \subset \mathbb{R}^n$, then the convex hull of X is the set of all convex combinations of its points, $H(X)$. In a convex combination, each point in is assigned a weight or coefficient in such a way that the coefficients are all non-negative and sum to one, and these weights are used to compute a weighted average of the points. For each choice of coefficients, the resulting convex combination is a point in the convex hull, and the whole convex hull can be formed by choosing coefficients in all possible ways. Expressing this as a single formula, the convex hull is the set:

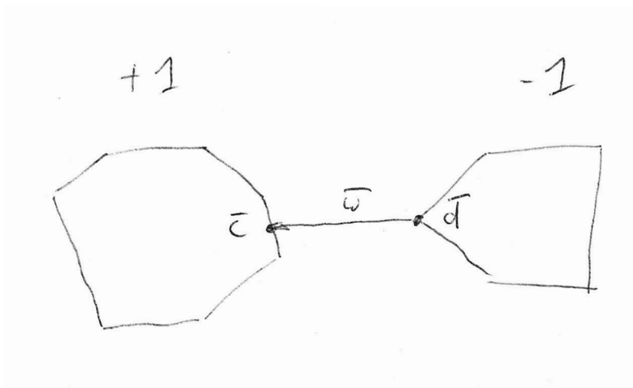
$$H(X) = \{\sum_{i=1}^{l} \alpha_i \overline{x}_i\}$$

with $\sum_{i=1}^{l} \alpha_i = 1$ and $\alpha_i \geq 0$.



SVM: The Separable Case

Let $D = \{(\overline{x}_1, y_1), \ldots, (\overline{x}_l, y_l)\} \subset \mathbb{R}^n \times \{+1, -1\}$ be our training data. Consider two class distributions +1 and -1 and their corresponding hulls $H(+1)$ and $H(-1)$,



We pick the point $\overline{c} \in H(+1)$ to be closest to the -1 class distribution and we pick point $\overline{d} \in H(-1)$ to be closest to the +1 distribution. Next we draw a vector from $\overline{d}$ to $\overline{c}$ such that

$$\overline{w} = \overline{c} - \overline{d}$$

Now, picking the points $\overline{c}$ and $\overline{d}$ as we did above and then drawing the vector $\overline{w}$ is the same as saying that we want to minimize the length of $\overline{w}$, in other words,

$$\min |\overline{w}| = \min \frac{1}{2}|\overline{w}|^2 = \min \frac{1}{2}\overline{w} \bullet \overline{w}$$

SVM: The Separable Case

Now consider that $\overline{c} \in H(+1)$ and $\overline{d} \in H(-1)$, therefore

$$\begin{aligned}
\overline{c} &= \sum_{\overline{x}_p \in +1} \alpha_p^+ \overline{x}_p \\
\overline{d} &= \sum_{\overline{x}_q \in -1} \alpha_q^- \overline{x}_q
\end{aligned}$$

Now, let $\overline{\alpha}$ be the *concatenation* of $\overline{\alpha}^+$ and $\overline{\alpha}^-$ with

$$|\overline{\alpha}| = |\overline{\alpha}^+| + |\overline{\alpha}^-| = l$$

then

$$\min_{\alpha} \frac{1}{2} \overline{w} \bullet \overline{w} = \min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \overline{x}_i \bullet \overline{x}_j$$

subject to

$$\begin{aligned}
\sum_{i=1}^{l} y_i \alpha_i &= 0 \\
\alpha_i &\geq 0
\end{aligned}$$

SVM: The Separable Case

It is worthwhile to take a look at the constraint

$$\sum_{i=1}^{l} y_i \alpha_i = 0$$

We can rewrite this constraint as

$$\begin{aligned}
\sum_{i=1}^{|\overline{\alpha}^+|}(+1)\alpha_i^+ + \sum_{i=1}^{|\overline{\alpha}^-|}(-1)\alpha_i^- \quad &= \quad \sum_{i=1}^{|\overline{\alpha}^+|}\alpha_i^+ - \sum_{i=1}^{|\overline{\alpha}^-|}\alpha_i^- \\
&= \quad 1 - 1 \\
&= \quad 0 \text{ (if the points fulfill the convex hull property)}
\end{aligned}$$

SVM: The Separable Case

Finally, we have

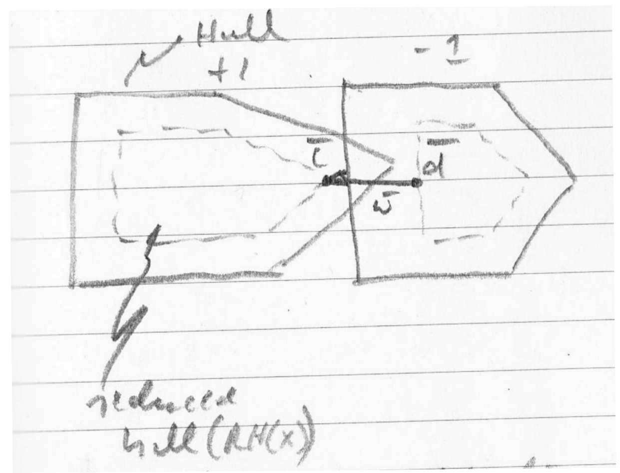
$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \overline{x}_i \bullet \overline{x}_j$$

subject to

$$\begin{aligned} \sum_{i=1}^{l} y_i \alpha_i &= 0 \\ \alpha_i &\geq 0 \end{aligned}$$

It is interesting to note that this looks very similar to the optimization problem that we derived via Lagrangian optimization theory.

SVM: The Non-Separable Case



The *reduced hull* $RH(X)$ is defined as

$$RH(X) = \{\sum_{i=1}^{l} \alpha_i \overline{x}_i\}$$

with

$$\sum_{i=1}^{l} \alpha_i = 1$$

$$C \geq \alpha_i \geq 0$$

SVM: The Non-Separable Case

The optimization problem then becomes

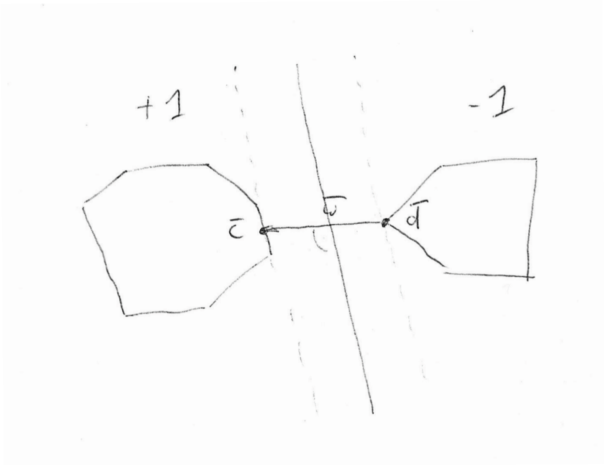
$$\max_{\alpha} - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j \overline{x}_i \bullet \overline{x}_j$$

subject to

$$\sum_{i=1}^{l} y_i \alpha_i = 0$$

$$C \geq \alpha_i \geq 0$$

Model



It is easy to show that our model is a support vector machine,

$$\hat{f}(\overline{x}) = \mathrm{sign}(\overline{w}^* \bullet \overline{x} - b^*)$$

with

$$\overline{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \text{ (think } \overline{w} = \overline{c} - \overline{d}\text{)}$$

and

$$b^* = \sum_{i=1}^{l} \alpha_i^* y_i \overline{x}_i \bullet \overline{x}_{sv+} - 1$$

Reading

C. Bennet – "SVM - Hype or Hallelujah" – available on the course website.